

ITS Analyzers

Meta-Analysis research on the effectiveness of Intelligent Tutoring System (ITS) products.

Vinay Rao (vrao41@gatech.edu), Gouri Guin (gguin3@gatech.edu)

CS6460 -Fall 2016, Georgia Institute of Technology, USA - Date: 12/06/2016

Abstract. Evaluation of intelligent tutoring systems (ITS) is an important area of research in current educational practices. There are many work has been done to analyse the effectiveness of various ITS products to aid student education. We focused on top 3 ITS products to compare and analyse individual ITS product outcome and its effectiveness. Additionally, we explored the current problems that may exist with measuring the effectiveness of ITS by conducting research study on different methods that are currently used.

Keywords: Intelligent Tutoring Systems, Computer-Assisted Learning, Meta-Analysis

1 Introduction/Motivation

Aid children education and to help eliminate student learning difficulties and help build confidence for continued education with the help of ITS adoption either through classroom integration or through ITS support in addition to classroom teachings was the main motivation for our research. Our research questions included:

- Can ITS effectively be integrated into student's education?
- Which module(s) within ITS is adding value the student education?
- What method best suited for measuring ITS effectiveness?

2 Intelligent Tutoring System (ITS)

Intelligent Tutoring System is a broad term which mainly compose any computer program that contains some intelligence and can be used in learning. Modelling an ITS has been around for four decades with many deep researches and some successful products specific to some study area. The primary goal of an ITS is to provide customizable feedback to student and adapt to student learning abilities mimicking human tutor to

adfa, p. 1, 2011.

© Springer-Verlag Berlin Heidelberg 2011

core. Many ITS systems are able to interpret complex student responses and can learn as they operate, they are able to discern where and why a student's understanding has gone astray and to offer hints to help the student understand the material at hand. There are studies to use ITS in addition to classroom teaching and also studies to adopt ITS instead of human tutors within classroom. [1]

In building an intelligent tutoring system, researchers quest for robust techniques and suitable approaches so that intelligent and interactive communications between learners and the system can be seamlessly established. ITS projects can vary tremendously according to the relative level of intelligence of the components.

However, although ITS has been a great success in targeted teaching, it has not been able to penetrate to wide array of field. This is mainly due to complexity involved in designing and developing ITS application, cost associated with development issues like assessing each individual student differently and adapting to student's mental model be it learning new language or teaching Mathematics. There are many research in this area like "Language Learning: Challenges for Intelligent Tutoring Systems" by Michael Heilman and Maxine Eskenazi [2] seems to highlight ITS challenges.

2.1 Products

In our research we chose to include many of the popular ITS applications like Cognitive Tutor (initially developed at Carnegie Mellon University) [3], AutoTutor by Memphis University [4], The Andes Physics Tutor (developed at Arizona State University) [5], W-Pal [6], BASIC Instructional Program (BIP) [7] and AutoTutorLite (Wolfe, Fisher, et al., 2012) [8].

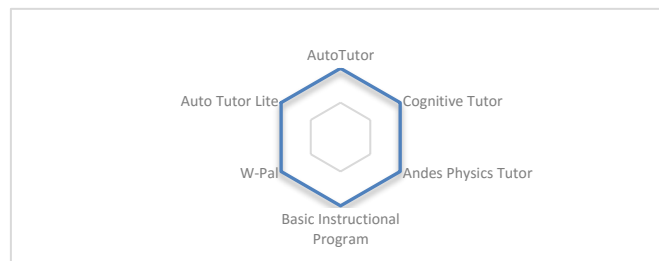


Fig. 1.

Based on the criteria we developed and also due to availability of past extensive research we zeroed on top 3 ITS products namely Cognitive Tutor, Andres Physics Tutor and AutoTutor.

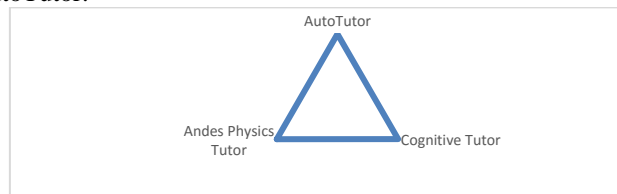


Fig. 2.

Overall, we focused on 29 evaluations comprised of many studies on various students (19,290) with varied effectiveness utilizing different meta-analysis method to measure their effectiveness [9], [10],...,[28].The methods we generally found within these evaluations were similar in nature. For the purpose of rating and ranking we developed seven categories of effectiveness (High Positive, Medium Positive, Low Positive, No Change, Low Negative, Medium Negative and High negative). We plotted our gathered research data to look at the distribution of effectiveness.

2.2 Cumulative ITS Effectiveness

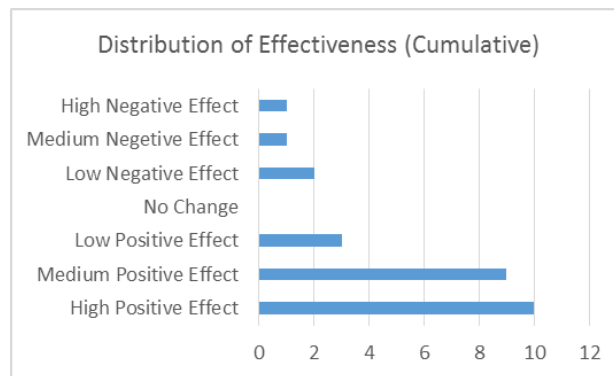


Fig. 3.

Overall, as expected most of the studies pointed at positive effect to the student's learning. As shown above many had High positive effect on student's learning which is very encouraging. However, there are instances where these studies pointed at negative effect this could be an outlier and needs further exploration.

2.3 ITS Product Effectiveness (%)

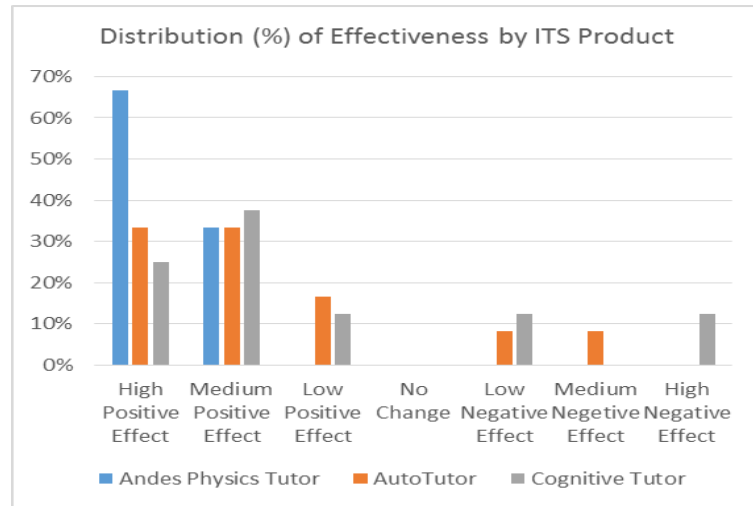


Fig. 4.

When compared the effectiveness across ITS products (Fig: 4), it was interesting to see Andes Physics Tutor outperformed other ITS products and had no negative effect during our chosen studies.

2.4 Modules

One of the first architectures of an ITS system was presented by Burn and Caps in 1988. This traditional ITS architecture was based on four modules: the domain model, the student model, the teaching model (pedagogical module), and a learning environment or user interface module as shown in diagram below.

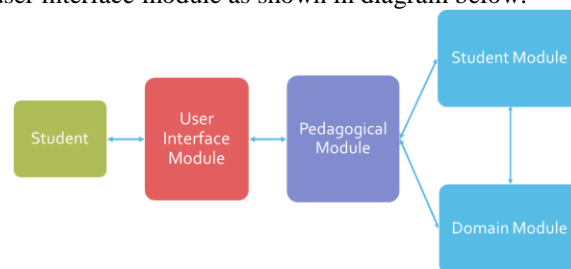


Fig. 5.

During our study the top 3 ITS products at a high-level adopted the same traditional architecture with some variations.

- Andes Physics Tutoring System has a very rich domain module to define knowledge effectively and has also very rich pedagogical module with many intelligent help,

feedback and hints. However it has a very mediocre paper-like UI and needs little learning from the user. It was interesting to note that it does not have student module and it was replaced by actual class room. Overall, it not customized for individual student but is setup for whole class.

- AutoTutor is an intelligent tutoring system that holds conversations with the human in natural language and is most modified ITS product in the market with many variations like Auto Tutor lite, etc. AutoTutor has produced learning gains across multiple domains (e.g., computer literacy, physics, critical thinking). Three main research areas are central to AutoTutor: human-inspired tutoring strategies, pedagogical agents, and technology that supports natural language tutoring.
- Cognitive Tutor utilized rich UI and adopted John Anderson's ACT-R theory of human cognition to provide immediate instruction in a problem-solving context. Its' domain module was built in such a way that it can be customizable and be implemented into classrooms as a part of blended learning that combines textbook and software activities.

	Andes Physics	Cognitive Tutor	Auto Tutor
Domain Module	Good	Moderate	Good
Teaching/Pedagogical Module	Good	Moderate	Good
Student Module	N/A	Good	Good
UI/Environmental	Moderate	Good	Good

Table 1.

Looking at the above table and based on the effectiveness analysis, it was evident that the ITS adopting a good teaching/pedagogical Module tends to be more effective and is probably one of the most important modules within ITS. It was interesting that Auto Tutor chose to adopt Latent Semantic Analysis (LSA) [32], which is a high-dimensional, statistic technique that measures conceptual similarity of any 2 pieces of text. Auto Tutor also adopted Dialogue management Module which brings in Natural Language Processing (NLP) by providing flexibility to adopt to the student by virtue of student ability (LSA score), student verbosity (number of words per turn), evolution of question's answer in dialogue history. Overall, Auto Tutor is built as a conversation partner to assist the learner accessing learning objects, using them and monitoring progress of learning objectives.

It was interesting that Andes Physics Tutoring system adopted classroom integration to deliver their ITS instead of dedicated Student module. This makes one wonder that a classroom adoption reinforces student's learning better than the individual setting.

Although, the UI module was weak in Andes Physics Tutoring system, the growing trends of mobile, game etc. reinforces the need to have a great UI which improves overall usability and also reduces the training required by the student to adopt ITS. The effective UI module may be the key in overcoming the ITS difficulty to penetrate in early education and is also an important module within ITS.

Finally, the domain module which is the brain for the ITS which holds the domain specific knowledge is the most important module within ITS. It was consistently rated good across all products, we evaluated. It is the key to have an accurate, in-depth and effective subject knowledge which is the important aspect to reinforce student's learning and help fill the gaps.

3 Meta-Analysis

Meta-analysis is a statistical technique, or set of statistical techniques, for summarizing and reviewing the results of previous quantitative research mostly combining several studies into a single estimate.

Beecher (1995) [29] undertook the earliest example of a meta-analysis and Glass (1976) [30] coined the term "meta-analysis" to refer to a philosophy, not a statistical technique. The meta-analysis method began as a statistical procedure for combining and comparing research findings from different studies focusing on similar phenomena (Nijkamp and Pepping 1997-98) [31], and a variety of meta-analytical methods have been developed in the past decades.

Since the focus of most of the research we evaluated was to measure the effectiveness of ITS product, the meta-analysis method "comparison study" was widely used across all top 3 ITS products. To complement the comparison study many other methods like survey, interview, Feedback were effectively adopted. The feedback/instruction quality was key in these studies as the opinion of SME in the area of teaching was valued very highly and the underlying goal of ITS product was to mimic human tutoring.

3.1 Methods

Evaluation of intelligent tutoring systems (ITS) is an important area of research. Based on the research paper we identified following meta-analysis methods were highly utilized.

1. Feedback/Instruction Quality
2. Product Evaluation
3. Expert Knowledge
4. Observation & Qualitative Classification
5. Comparison Studies

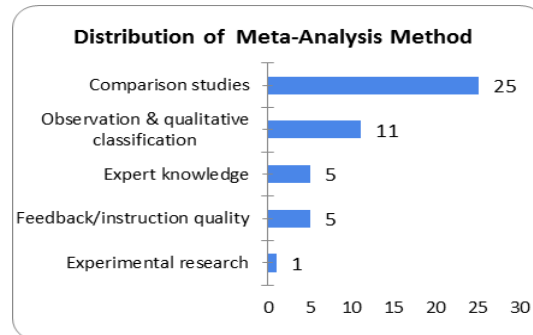


Fig. 6.

The biggest gap in meta-analysis study we found was no consistency in terms of technique adopted to measure the effectiveness of ITS. Also, a meta-analysis of several small studies does not predict the results of a single large study. The researcher's passion towards the field might bring bias that cannot be controlled by the meta-analysis method. A good meta-analysis of badly designed studies may also result in bad statistics. It is evident that there is strong correlation between meta-analysis and study itself. As a result, we first want to focus on existing Intelligent Tutoring Systems and look at their studies. As mentioned before since we are depending on past studies we are concerned about lack of information regarding bias associated with methods. Additionally, methodological selection criteria introduce unwanted subjectivity, defeating the purpose of the approach.

It is very important to note that our top performing ITS product Andes Physics Tutoring System not only adopted both hourly and final exams but repeated exams over five year period conducting multiple studies as oppose single exam adopted in Auto Tutor and Feedback technique adopted in Cognitive Tutor along with repeated 2 year studies. Even in Cognitive Tutor the second year results in many cases were much encouraging than the first year results.

During our research both Andes Physics Tutoring System and Auto Tutor which outperformed Cognitive Tutor conducted exams/tests to measure the effectiveness of ITS, whereas Cognitive Tutor heavily depended on techniques like interviews and surveys.

This makes us to believe that an effective way to measure the effectiveness of ITS products can be more accurate when the tests conducted repetitively as there is an additional learning curve for students on top of the subject being thought which is to adopt to ITS product itself. This conclusions was further substantiated by the Andes Physics tutoring study where the first year effectiveness was comparatively lower as oppose to consecutive year test.

The Figure 7 shows the distribution of the different methods adopted at more granular level. And Figure 8 aggregates those to the methods in Industrial standard terms.

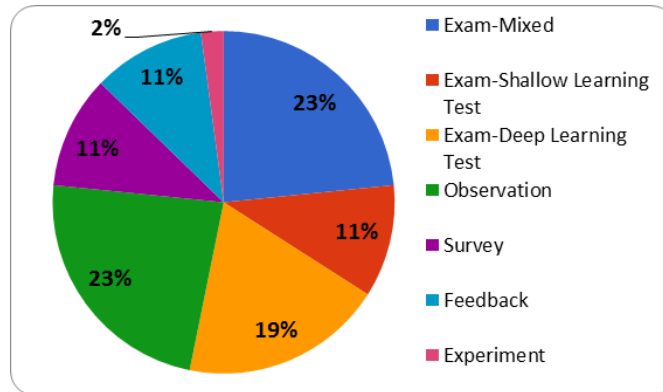


Fig. 7.

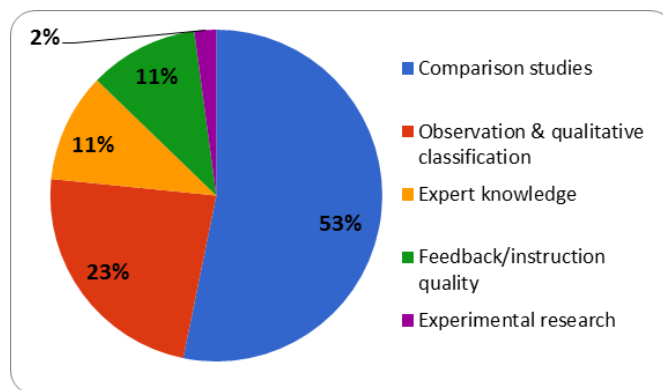


Fig. 8.

4 Proposal

Based on our research of 29 past ITS product evaluations comprised of many studies on various students (19,290) with varied effectiveness utilizing different meta-analysis method to measure their effectiveness, we found that industry needs to standardize the meta-analysis framework.

The following chart shows the target audience for our studies where the studies were targeted to students of different grades with most of the studies (~70%) targeted to college students. It was clear that ITS is more effective when measured with students going to college as oppose to students in going to K-7 to K-12. This might reinforce the fact that ITS needs some learning from student and higher the grades higher chances of successful ITS adoption.

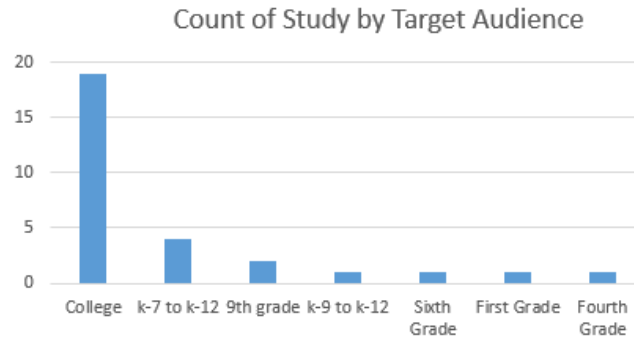


Fig. 9.

Meta-analysis has now become a widely accepted research tool, encompassing a range of procedures used in a variety of disciplines. Various meta-analysis frameworks have been brought out in last three decades. Although it is widely accepted that with the basic principles, the techniques have been widely different according to characteristics of research fields. Under the Meta-analysis framework, appropriate methods can be selected according to different research questions.

One goal of a meta-analysis will often be to estimate the overall, or combined effect of ITS product. Based on the studies we reviewed we were unable to simply compute the mean of the effect sizes as studies adopted different meta-analysis technique with some studies more precise than others. It would have been nice to adopt some kind of weight based on study and its preciseness.

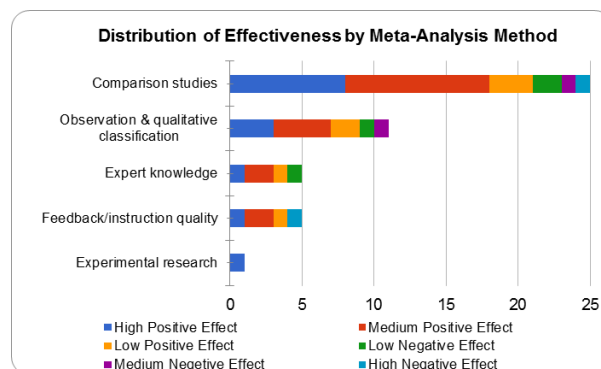


Fig. 10.

We recommend adopting a weight based on study and meta-analysis method adopted. For example: exam/test have higher precision over survey, observation or feedback. Thus Quantitative analysis technique like exam/test would have higher weightage compared to other qualitative analysis techniques. Also, it apparent that Qualitative analysis is less reliable because it is more subjective than quantitative methods, and may generally yield smaller sample sizes.

In our research scope if we adopt such weight then Andes Physics Tutoring System would have higher combined effect size thus would be out-performing Auto Tutor and Cognitive Tutor even higher, as Cognitive Tutor adopted both Quantitative and Qualitative analysis in their study to measure the effectiveness of ITS. On the contrary if Cognitive Tutor would have adopted our proposed weight for effect size measurement then it probably would have showed better results provided the test/exam had better outcome which we were not able to verify as we only had access to combined effect size.

Additionally, we propose comprehensive mixed method by effectively combining quantitative and qualitative studies together as they sometime address multiple forms of evidence with higher weightage to Quantitative as shown pictorially below.

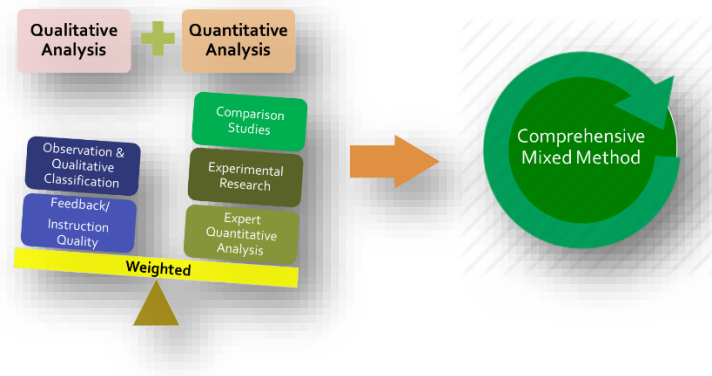


Fig. 11.

When we looked at the distribution of effectiveness by meta-analysis as shown below, we felt that an effective mixed method combining Comparison Studies, Observation & Qualitative Analysis with Experimental research would show better result as they had higher positive impact in our selected studies.

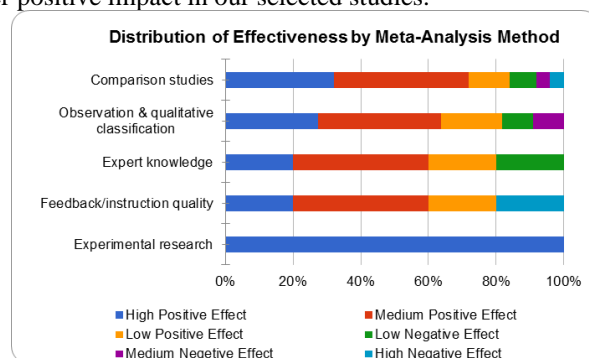


Fig. 12.

Our proposal within ITS module space is to build an ITS with all four modules as listed below:

- Mobile Friendly User Interface Module
- NLP based pedagogical Module
- Artificial intelligence/ cognitive science backed Domain Module (Example: integrating IBM Watson to domain module).
- Customizable Student Module, which can adopt to both classroom settings and individual settings.

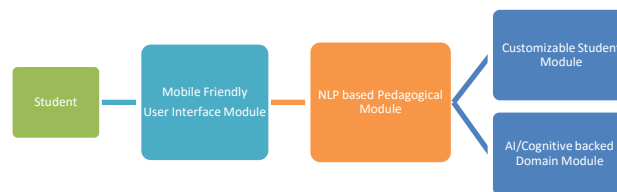


Fig. 13.

5 Conclusions

From our research it was evident that ITS brings novel and useful techniques to enhance students learning and can effectively be integrated into student's education. We believe our proposed changes to four core ITS modules as shown in figure 13 will make ITS more effective in education.

As our next step, we are interested in partnering with some educational institute to develop an ITS with changes to modules as suggested above. With more intelligent modules (backed by AI, Cognition, Natural language Interaction and Machine Learning) we feel ITS can make an impact to students across all grades. We will be interested to see if it makes any positive impact in early child education.

We also want to partner with teachers and other education researchers to carry out largescale user trials with our proposed meta-analysis framework as shown in figure 11 to measure effectiveness of our top 3 ITS product along with our newly proposed ITS solution in the classroom setup and also individual setup. Data from such trials will generate new insights into the effect of ITS on student's ability to learn in various domains.

6 Acknowledgments

We thank our Professor David Joyner and our research mentor Ken Brooks at Georgia Institute of Technology College of Engineering for their guidance and support.

7 References

1. Koedinger, K., & Tanner, M. (n.d.). 7 Things You Should Know About Intelligent Tutoring Systems. Retrieved July 09, 2013, from
2. Heilman, M., & Maxine, M. (n.d.). Language Learning: Challenges for Intelligent Tutoring Systems. Retrieved from http://www.learnlab.org/uploads/mypslc/publications/heilman_eskenazi_its2006.pdf
3. Ritter, S., Anderson, J.R., Koedinger, K.R., Corbett, A.T. 2007. Cognitive Tutor: applied research in mathematics education. *Psychon. B. Rev.* 14 (2007), 249-255.
4. AutoTutor. (n.d.). Retrieved September 24, 2007, from <http://www.memphis.edu/iis/projects/autotutor.php>
5. Kurt VanLehn, Brett Van De Sande, Robert Shelby, Sophia Gershman - The Andes physics tutoring system: An experiment in freedom
6. Rod D. Roscoe, Laura K. Allen, Jennifer L. Weston, Scott A. Crossley, Danielle S. McNamara, The Writing Pal Intelligent Tutoring System: Usability Testing and Development
7. Barr, A., beard, M., & Atkinson, R. C. (1975) A rationale and description of CAI program to teach BASIC programming language. *Instructional Science*
8. Wolfe, C. R., Widmer, C. L., Reyna, V. F., Hu, X., Cedillos, E. M., Fisher, C. R., Weil, A. M. (2013). The Development and Analysis of Tutorial Dialogues in AutoTutor Lite. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3748232/>
9. Cabalo, J. V., Jaciw, A., & Vu, M.-T. (2007). Comparative effectiveness of Carnegie learning's Cognitive Tutor Algebra I curriculum: A report of a randomized experiment in the Maui School District. Palo Alto, CA: Empirical Education, Inc.
10. Ritter, S., Kulikowich, J., Lei, P., McGuire, C., & Morgan, P. (2007) what evidence matters? A randomized field trial of Cognitive Tutor® Algebra I.
11. Agodini, R., M. Dynarski, M. Honey, and D. Levin. "The Effectiveness of Educational Software: Issues and Recommendations for the National Study." Report prepared for Institute of Education Sciences, U.S. Department of Education. Princeton, NJ: Mathematica Policy Research, Inc., May 2003
12. John F. Pane, Beth Ann Griffin, Daniel F. McCaffrey, Rita Karam (2014), Effectiveness of Cognitive Tutor Algebra I at Scale.
13. Shneyderman, A. (2001). Evaluation of the Cognitive Tutor Algebra I Program. Miami, FL: Miami-Dade County Public Schools Office of Evaluation and Research.
14. Bibi, Tauqir, "Analysis of Cognitive Tutor Geometry Curriculum" (2010). Graduate Theses and Dissertations. Paper 11554.
15. Graesser, A. C., Jackson, G. T., Mathews, E. C., Mitchell, H. H., Olney, A., Ventura, M., & Chipman, P. (2003). Why/AutoTutor: A test of learning gains from a physics tutor with natural language dialog. In R. Alterman & D. Hirsh (Eds.), *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society* (pp. 1–6).
16. Heilman, M., & Maxine, M. (n.d.). Language Learning: Challenges for Intelligent Tutoring Systems. Retrieved from http://www.learnlab.org/uploads/mypslc/publications/heilman_eskenazi_its2006.pdf
17. Allen, J. (1995). *Natural language understanding*. Redwood City, CA: Benjamin/Cummings.
18. Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, 4, 167-207.
19. Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of educational Psychology*, 94, 416-427.

20. Person, N. K., Graesser, A. C., Bautista, L., Mathews, E. C., & the Tutoring Research Group (2001). Evaluating student learning gains in two versions of AutoTutor. In J. D. Moore, C. L. Redfield, & W. L. Johnson (Eds.), *Artificial intelligence in education: AI-ED in the wired and wireless future* (pp. 286-293). Amsterdam: IOS Press.
21. Person, N. K., Graesser, A. C., Kreuz, R. J., Pomeroy, V., & the Tutoring Research Group (2001). Simulating human tutor dialogue moves in AutoTutor. *International Journal of Artificial Intelligence in Education*, 12, 23-39.
22. Aleven, V., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science*, 26, 147-179.
23. Allen, J. (1995). *Natural language understanding*. Redwood City, CA: Benjamin/Cummings.
24. Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, 4, 167-207.
25. Albacete, P. L., & VanLehn, K. (2000a). The Conceptual Helper: An intelligent tutoring system for teaching fundamental physics concepts. In G. Gauthier, C. Frasson & K. VanLehn (Eds.) *Intelligent Tutoring Systems: 5th International Conference, ITS 2000* (pp. 564-573). Berlin: Springer.
26. Albacete, P. L., & VanLehn, K. (2000b). Evaluation the effectiveness of a cognitive tutor for fundamental physics concepts. In L. R. Gleitman & A. K. Joshi (Eds.) *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society* (pp. 25-30). Mahwah, NJ: Erlbaum.
27. Kurt Vanlehn , Collin Lynch , Kay Schulze , Joel A. Shapiro , Robert Shelby , Linwood Taylor , Don Treacy , Anders Weinstein , Mary Wintersgill, The Andes Physics Tutoring System: Lessons Learned, *International Journal of Artificial Intelligence in Education*, v.15 n.3, p.147-204, August 2005
28. S. W. Bonham, D. L. Deardorff, and R. J. Beichner, "Comparison of student performance using web and paper-based homework in college-level physics," *Journal of Research in Science Teaching*, vol. 40, pp. 1050-1071, 2003
29. Beecher HK. The powerful placebo. *JAMA* 1955;159:1602 –1606
30. Glass GV. Primary, secondary and meta-analysis of research. *Educ Res* 1976; 10:3 – 8.
31. Bergh, J.C.J.M., van den, K Button, P Nijkamp and G.Pepping (1997), *Meta-analysis in Environmental Economics*, Dordrecht: Kulwer.
32. Lisa M. PytlikZillig Patall, E. A. (n.d.).- *Technology-based Education: Bringing Researchers and Practitioners Together*
33. Comprehensive Meta-Analysis., from https://www.meta-analysis.com/pages/why_do.php
34. 19.5 ITS EVALUATIONS. (n.d.). Retrieved August/September, 2001, from <http://www.aect.org/edtech/ed1/19/19-05.html>
35. Attali, Yigal, & Burstein, Jill. (2006). Automated essay scoring with e-rater V.2. *Journal of Technology, Learning, and Assessment*, 4(3). Retrieved <http://files.eric.ed.gov/fulltext/EJ843852.pdf>
36. Barab, Sasha, Thomas, Michael, Dodge, Tyler, Carteaux, Robert, & Tuzun, Hakan. (2005). Making learning fun: Quest Atlantis, a game without guns. *Educational Technology Research and Development*, 53(1), 86–107.
37. Bell, Courtney, & McNamara, Danielle S. (2007). Integrating iSTART into a high school curriculum. *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (pp. 809–814). Austin, TX: Cognitive Science Society.
38. Duit, R., and Treagust, D. F. Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education* 25, 6 (2003), 671–688

